Drawing graphs of causal relationships

2026-02-27

Causation in my household

It helps to start with an example. The other day I walked into the kitchen and found a puddle of water on the counter. Being a good scientist, in addition to an annoyed parent, I started to consider what caused this puddle to be there. One idea immediately leaped to mind: one of my kids had probably grabbed some ice for their drink and left a few pieces on the counter, which melted and formed a puddle.

Now how to draw this? That is, after all, the reason for this story.

Enter the DAG

We use graphs and diagrams a *lot* in science, to show relationships in ideas or data, highlight comparisons, tell stories, and, as in this case, illustrate how we think the world, or a small little bit of it, works. In this case, we will use a box-and-arrow sort of graph to describe *what causes what* in my kitchen[1].

[1] Other examples in science might be more compelling, sure, but they always say to write what you know!

First, a couple of rules:

1. The boxes will represent the things that are playing a part in this causal story. They are things or variables or

whatnot. In general they are quantities we could at least theoretically measure[2]

[2] They are *not* processes, like "melting" or "heating".

2. The arrows will represent *causal* relationships such that $x \to y$ implies that a change in x *causes* a change in y. Note that we are not drawing things in order, what comes first or second, but rather cause and effect relationships[3].

[3] And could represent particular processes, like "melting."

3. The arrows do not imply a direction (positive or negative) or shape (linear, sigmoidal) to the causal relationship, just that there is one.
4. We do not draw interactions. If this means nothing to you, fine, just move along to number 5. If it does, just note that diagrams with additive or interactive or contingent relationships all look the same.
5. No cycles. That is, you should not be able to follow the arrows to get back to where you started from.

Here is what my initial graph might look like (Figure 1).

Figure 1: A simple DAG.

This is a directed (meaning arrows) acyclic (meaning no cycles) graph (nodes and edges), or a DAG. Don't worry about the jargon, just note that they are a simple, clear way of describing how we think this system works[4].

[4] They are also super handy in the practice of science and statistics, helping us understand which statistical models to use and what parameter estimates represent. You'll see them in epidemiology, econometrics, sociology, and increasingly in ecological research. If you stay in science, you'll see them. You can thank me later for introducing you to DAGs.

In this case we would read this as, "my kid caused ice to be on the counter" and then "the ice caused water to be on the counter".

But wait, there is probably at least a bit more. Ice doesn't turn to water without melting, right? That requires heat! So we need to add "heat" to the diagram.

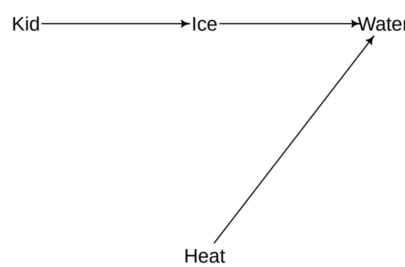

Figure 2: A less simple DAG.

So our diagram (Figure 2) now adds, "a change in heat causes a change in water," which sounds weird because this only makes sense when there is ice present, too. But our diagram is about the system (in my kitchen) as a whole. This arrow does not necessarily say that heat causes water *in general.* It just says that if we were to change the heat (e.g., remove it from the room) we would change the presence of liquid water on the counter. Note, however, that the graph does not *by itself* imply that you need both ice and heat to get water. It just sidesteps the whole issue. *We*, as thinking people, add that interpretation[5]. So now we have arrows saying that if we were to change the

[5] See rule 4, above, about "interactions." This is what I meant.

ice on the counter or the heat in the room we would cause a change in the water on the counter.

If your head hurts, that's probably a good sign. It suggests you are paying attention. This is not always an intuitive way of describing a system or drawing a graph, at least not at first.

Anyway, we might need to complicate our diagram a bit, because I suspected that my kid wanted ice because it was hot out. In other words, it seemed like heat in the house on this warm day *caused* the ice to be on the counter, along with my sloppy kid. Thus this graph to the right

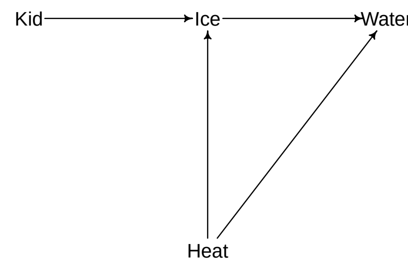

Figure 3: An even less simple DAG.

What this graph (Figure 3) means is that a change in heat causes a change in water on my kitchen counter through two **causal paths**: 1) by causing the ice to be present or not, which then can turn into water ($Heat \rightarrow Ice \rightarrow Water$) and 2) by providing or not providing the energy to melt the ice ($Heat \rightarrow Water$). Having shapes like this is fine in these graphs just so long as the arrows do not form a cycle. We could even add an arrow for the causal effect of, say, my cat who likes to play with ice cubes. Where would you add this causal relationship?

The point?

What, you may be asking, is the point of this? Well, setting aside all of the wonderfully useful logical and statistical magic this enables[6], which we won't be using[7], these diagrams force us to think very carefully about causation. They also help us think through "what if" scenarios, like "what if we added or removed the heat?" or "what if we added another kid?" That is, they help us see the consequences of **interventions**, which might include experimental manipulations or public health measures or anything else we might try. For instance, this diagram suggests that making sure there are ample paper towels would not change anything because we have no way for them to cause any changes[8].

[6] see the footnote, above

[7] Phew!

[8] But we could create a different DAG with toweling in it.

The other point to highlight is that this diagram is, in essence, a hypothesis. It both describes the hypothesis, the statement about how the world (or at least my kitchen) works, and what would change given particular interventions. But if this model is wrong, then so are its predictions about interventions.

Again, you might be asking, "well what's the point" if it can be wrong. I have two responses. First, welcome to the real world where there are no magic wands[9] that can tell us the Truth with a capital "T." Second, as I will emphasize in this class, we simply cannot make sense of *anything* in the world without invoking some sort of model. We do this instinctively, telling stories, but drawing out a DAG forces you to be clearer and makes it easier for someone else (sometimes *you*) to see how you are thinking about the world. So yeah, the model can be wrong, but that's not the model's fault. *You* made the model! Also, stop worrying so much about being wrong. It's unhealthy!

[9] Or statistical or other tools.

An alternative

There is absolutely nothing wrong with having an alternative working model in mind. In fact it is encouraged! So let's consider an alternative explanation and DAG for the water on the counter: it's rain water. In this scenario I'm thinking that the kid left the window open which let in the rain which then pooled on the counter. Like this one to the right (Figure 4).

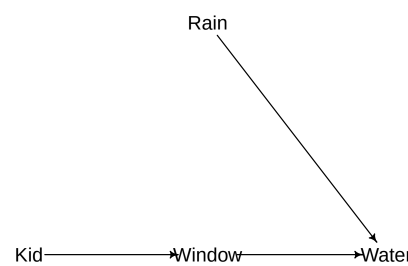

Figure 4: An alternate DAG.

Did that match your expectations? Oh, and we can add in heat, too, because heat would cause the water to evaporate, causing it to *not* be on the counter. Oh and maybe the kid opened the window because it was hot in the kitchen, so there might be an arrow there, too (Figure 5).

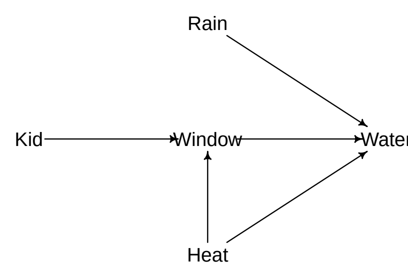

Figure 5: An alternate DAG, but now with heat!.

Before ending this silly example I would like to point out two related, perhaps not obvious consequence of this last DAG (Figure 5): 1) It suggests that if we knew the status of the window, whether it was open or closed, knowing whether the kid was in the kitchen opening or closing anything provides no extra information about whether we should see water on the counter. This will make more sense in larger studies with more data, but if you have data on the consequence (or **descendant**) of some **upstream** cause, then information on that upstream cause does not help you predict anything or describe anything in the data.

It's more of a statsy thing, but it may crop up in our work in this class. 2) If we were intervene by, say, nailing the window open or shut, we remove the influence of the kid on water on the counter, and of heat on the status of the window. Indeed, that's what experiments and interventions do: they remove dependencies in our system. In that case, our DAG would look like this last one on the right (Figure 6).

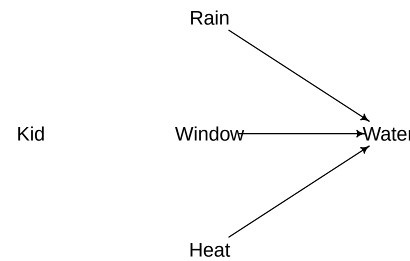

Figure 6: An alternate DAG, simplified.